Accounting for water molecules in drug design

S. E. Wong, F. C. Lightstone

August 20, 2010

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Title: Accounting for water molecules in drug design: A Review

Authors: Sergio Wong and Felice Lightstone

Abstract

Importance of the field

Areas covered in this review

What the reader will gain

Take home message

Paper length = 4000-6000 words

1. Introduction.

Achieving high affinity for a protein target is a necessary and essential requirement for drug design. While formally analytical expressions, based on statistical mechanics, exist that calculate binding free energies, it is highly difficult to compute binding free energies in practice[1].   It is non-trival to capture solvation effects, especially the effect is due to individual water molecules rather than the bulk properties of water.   These granular effects are unavoidable; a survey of structures in the PDB shows each ligand-protein complex has about 4.6 ligand-bound waters[2].   Furthermore, water not only stabilizes ligand interactions but plays a biological role in dictating specificity in [REF].

It is not surprising that modeling water raises such complications. The first issue is chosing an adequate representation. There are continuum solvent models where water is represented as homogeneous dielectric[3-6]. Other models include explicit water molecules[7-10]. Even more sophisticated models have a quantum representation of water molecules.   Water is modeled for industrial applications over a broad range of temperatures and pressures, but even quantum mechanical treatments of water have difficulty reproducing all properties of water[11-12].

In drug design water complicates the process of finding poses and calculating binding affinity. At the hit identification stage, via virtual screening, crystal structures of the target can include several crystallographic waters and the modeller must decide which waters, if any, to retain for the docking calculation. A broader issue is that solvation is often neglected in docking scoring functions and that can skew scoring results. Once a lead or leads are identified, and x-ray structures are available for optimization, water molecules may mediate contacts between the ligand and protein target. In that case, it is unclear if displacing the water molecule(s) will improve or deteriorate binding affinity. These are two central questions to dealing with water molecules in drug design.

Whether displacing a given structural water will improve binding depends on the balance between the enthalpy of contacts of the water molecule and the entropy change due to transferring the water to bulk solvent.   Dunitz[13] estimated ~ 2kcal/mol net effect. However, the contribution of individual waters is highly dependent on the environment

There are examples where it helps, while in some cases it does not . Mikol et al[14] found that displacing a water molecule can lower binding affinity.

In this review we address how to deal with bulk solvent for docking calculations, the issue of which water molecules to keep in docking calculations and which waters to displace during lead optimization. The discussion starts with representations of water models.

2. Representations of water

2.1 Continuum dielectric representations

In classical electrostatics, it is possible to calculate the polar contribution to solvation for a charged species by representing the solvent as a continuum dielectric[15]. In the case of macromolecules, there are usually ions in solution that can be incorporated into the model via the Poisson-Boltzamann equation[16]:

$$(1) \quad \nabla^2 \Psi = \frac{zen_\infty}{D\varepsilon} \sinh\left(ze\Psi / kT\right)$$

Where ψ e represents the charge of an electron, z corresponds to the ionization state of each ion (for example, z=1 for Na^+ while z=2 for Mg^{2+}), n_∞ is the concentration of position ions when the potential is zero. D corresponds to the dielectric constant of the ionic solution, k represents boltzman's constant, and T is the temperature. This equation can be linearized, in the case of small electrostatic potential, to yield the linearized Poisson-Boltzmann equation:

$$(2) \quad \nabla^2 \Psi = \kappa^2 \Psi$$

Where;

$$(3) \quad \kappa = \frac{2z^2 e^2 n_\infty}{D\varepsilon_o kT}$$

Delphi[REF], APBS [REF], UHBD [REF], calculate numerical solutions to these equations for a given biomolecule. This is a completely appropriate treatment for bulk properties because the solvent is a polarizable dielectric. However, it can be computationally costly and thus motivated a more approximate, but much faster approach: the Generalized Born treatment.

Basically, the generalized Born treatment approximates the Poisson-Boltzmann equation with a fairly high accuracy (see for example Edinger et al[6] and Tsui et al[5] and references therein). The polar contribution to solvation free energy is given by[3]:

$$(4) \quad \Delta G = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon_w}\right)\sum_{i,j} \frac{q_i q_j}{f_{GB}}$$

Where qi and qj are partial charges, e is the dielectric constant for water and fGB is defined as:

$$(5) \quad f_{GB}(r_{ij}) = \left[r_{ij}^2 + R_i R_j \exp\left(\frac{-r_{ij}^2}{4R_i R_j}\right) \right]^{1/2}$$

Where rij is the interactomic distance between atoms i and j; Ri and Rj denote effective Born radii for their corresponding atoms. This approximation has roots Born's analytical expression for the energy of a charged ion in a dielectric[17]:

$$(6) \quad \Delta G_{solvation} = -\frac{q^2}{2r}\left(1 - \frac{1}{\varepsilon_w}\right)$$

The expression is not valid for an irregular shape. Given an irregular shape, the Born radius for an atom corresponds to the value of R that when plugged into equation 5 would yield the correct solvation energy.

Inhomogeneous continuum water model is a different way of calculating the polar contribution to solvation of ligands and proteins. It is based radial distribution functions of the solvent. What is different about this method from the methods above is that the radial distribution functions for a given protein are calculated via molecular dynamics simulations. [FURTHER DISCUSS]

2.2 Explicit water models

Bulk solvent representations may not capture granular effects of water e.g. water mediated contacts and hydrogen bonding directionality [refs]. To address these issues, it is possible to employ an explicit solvent model where the oxygen and hydrogen atoms are explicitly included. The TIP3P model has a partial charge on the oxygen atom, a partial charge on each hydrogen and a van der Waals center at the oxygen [REF]. Another 3 point model is the SPC and SPC/E models. While these three models use 3-points to describe water, they yield different diffusion coefficients and radial distribution functions[REF]. The TIP4P model also has charges at the atomic centers but the van der waals center is displaced from the oxygen by [FILL-IN]. TIP5P waters have charge points at the position of the oxygen lone pairs [REF].

3. **Water in virtual screening calculations**

3.1 **how much does including waters affect docking calculations?**

Often, preparing a protein for a virtual screening calculation includes removal of all crystallographic waters from the binding site. Exceptions are made if, for example, there is some evidence suggesting a water is tightly bound to the protein and is effectively part of the protein structure. Without this information, it is difficult to justify keeping one or a subset of water molecules. Docking into all permutations of water arrangements can lead to a large, and impractical, number of possibilities. This raises the question of the trade -off between accuracy and the level of sophistication employed to account for these water molecules in docking calculations. The concern is not merely for the virtual screening stage, but rather for the effect on the overall drug design effort.

Huang and Shoichet[18] found that removing all crystallographic waters did not deteriorate enrichment for 12 of 24 target proteins. In their work they switched waters on and off during the screening and compared enrichment results with the case where all waters were removed. Nissink et al[19] docked into two sets of proteins: 1) 55 protein structures with no mediating waters and 2) 40 protein structures with at least one mediating water. Docking into structures with no mediating waters yielded a success rate of 73% (1.5A ligand RMSD threshold). That is, in 73% of the cases, the docked ligand pose was within 1.5A RMSD of the known crystallographic pose. In cases where the crystal structure has mediating waters, but they were removed for this exercise, the success rate was 61% at the 1.5A ligand RMSD threshold. As the threshold increases the differences dissipate (81 vs 79% at 2.5A ligand RMSD threshold). In another study using GOLD, including key crystallographic waters in the docking against influenza virus neuraminidase did not improve the docking poses with respect to the same exercise without waters. In addition, Garcia-Sosa and Mancera[20] noticed that including a tightly bound water adds further restrictions on the scaffolds that fit in the binding site and effectively reduce the chemical diversity of the scaffolds found in a de novo design study. Naturally, that added restriction may be necessary if displacing the water will deter binding, but removing that restriction does not prevent other scaffolds that can meet this restriction from being found via docking.

Other studies show including key water molecules improves the docked poses. For example, Roberts and Mancera[21] found that including water in docking calculations improves the quality of docked poses. The success rate at a 2.0 A ligand RMSD threshold was 74.2% when including crystallographic waters versus 65.8% without them. 240 structures made up the test set. Furthermore, including key waters in cross-docking calculations, where only waters consistently observed among the protein structures were included, also showed improvement in the quality of poses when compared to the same calculation without water molecules[22]. The success rate at the 2.2A threshold almost doubled (38.40 vs 69.39%) due to including water molecules. de Graaf et al[23], in their docking studies against chrochrome P450 and thymidine also found deterioration of poses in the absence of key water molecules. Some of these effects may be circumstantial. Hritz et al[24] found that the effect of water was highly dependent on the conformation of cytochrome P450 2D6 and that it was not consistently an improvement. Including protein flexibility in their docking calculations yielded results just as good as those including water on a static calculation. Furthermore, using molecular dynamics simulations it was possible to capture an effect from the water on docking results, but the effect was not consistently an improvement over the water-less structure[25].

While it is intuitive to use dock poses to capture the effect on accuracy from structural water molecules, a more practical metric to use is enrichment factor. Unfortunately, that kind of study is far less common. The work by Huang and Schoichet[18] is an example showing what one might expect: in some cases there is no significant impact in enrichment of true binders. However, this remains an open question.

3.2 Implicit solvation in docking calculations

MM-GB/SA is a practical compromise between the higher accuracy of a Poisson-Boltzmann solver and a fast, efficient method to calculate solvation free energies. Shoicket et al found that introducing solvation correction via MM-GB/SA improved enrichment in a screen against thymidylate synthase, the L99A mutant of T4 lysozyme and dihydrofolate reductase (DHFR)[26]. Graves et al also employed MMGBSA as a rescoring metric to improve enrichment in docking calculations[27]. Taylor et al also employed MMGBSA in their validation of FDS[28]. Lee and Sun[29] used the GOLD 79 protein/ligand complex set to evaluate MMGBSA and found that 85% of the time the docked pose was within 2.0 A RMSD of the crystallographic structure. Lyne et al[30] is another study, this time studying kinase inhibitors, showing an improved correlation (r^2>0.7) between relative binding data and MMGBSA scores.

The accuracy of MMGBSA re-scoring can sometimes even be useful for lead optimization[31]

3.3 Incorporating water explicit waters in virtual screening

There are two broad approaches to dealing with waters in virtual screening calculations. One approach is to identify structural waters in the target protein structure and subsequently determine which water(s) to keep during the docking calculation. That is, the goal is to determine which water(s) to keep *a priori*. A second approach is to include or remove water molecules during the docking calculation. A discussion of both approaches follows.

Methods to identify hydration sites in protein binding site vary from those based on conservation across protein families to computationally expensive free energy calculations based on physics alone. Bottoms et al[32] developed a method to identify conserved solvent sites. Examination of 6 protein families yielded known and new solvent sites. All families included solvent sites with than 90% conservation. Other studies found conserved water molecules in T4 lysozyme[33], aspartic proteinases[34], Rossmann dinucleotide-binding domains[35], serine proteases[36], and legume lectins[37]. Ehrlich et al[38] employed a neural network to predict hydration sites based on protein sequence. From the protein sequence two neural networks predicted secondary structure and solvent accessibility. This information was used as input, in addition to sequence to a third neural network that predicted hydration sites with a success rate of 60-70%. Thus, highly conserved water molecules may be identified within a protein family.

Other efforts are based on more physical metrics. For example, Poornima and Dean[39] argued that the shape of the protein surface affects how tightly water molecules will bind the protein. Specifically, they proposed deep grooves and micro-grooves on the protein-site are best suited to tightly bind water molecules. Goodford[40] developed a program (GRID) to identify favorable interaction sites for probe molecules. One of these probes is water. Interactions were computed based on Lennard Jones, electrostatics and hydrogen bonding. The Hydropathic Interactions (HINT) "natural" force field was developed to evaluate the contribution of water in protein-ligand interactions[41]. It employs a pairwise potential based on non-covalent interactions: electrostatic, acid-base, hydrophobic and hydrogen bonds and experimental values of logP (octanol/water partition) and can be used to evaluate ligand binding

energies. The advantage of this approach is that it provides more meaningful/intuitive, perhaps "natural", interpretation of the results. A case study of waters in HIV-1 protease showed that including a key water molecule yielded a correlation coefficient ~ 0.6 with experimental binding data.

Michael et al[42] were able to accurately find the correct number and position of water sites in 5 binding sites using their "Just Add Water Molecules" (JAWS) procedure.   JAWS uses a double-decoupling approach to compare the removal of a water molecule from the bulk and from the binding site.   A unique feature of this method is the use of a scaling parameter θ that modulates intermolecular interactions for a given water molecule.   Each θ is considered a degree of freedom in the monte-carlo calculation. This approach can also yield the free energy of transferring a water molecule from its binding site cavity to bulk solvent. However, the free energy calculations can suffer from poor convergence.

Once hydration sites are known, it is key to determine which water molecules are displaceable. Approximate methods are able to identify displaceable waters given structural waters. *Consolv* is based on a K-nearest neighbor genetic algorithm that takes as input B-factors, number of hydrogen bonds, hydrophilicity and atomic density. After training on 13 proteins and 2334 first-shell waters, it yields 61% accuracy[43] in indentifying structural waters. The authors note that the microenvironment of the water is the key determinant of water affinity. Amadasi et al[44] report being able to identify "relevant" waters in 86% of the cases examined using the HINT "natural" force field[45]. Relevant waters were defined as those waters generally conserved but that may be displaced during the lead optimization stage. Waterscore[46] is a metric based on a multivariate analysis of B-factors, solvent contact surface area and the number of protein atomic contacts (NPAC) that identifies structural waters displaced by ligands. It yielded 67% accuracy using a 1.5 A threshold for identifying displaced waters. These methods are amenable for initial docking calculations given that they do not require massive computations, such as in the case of free energy calculations, to yield an answer and have an acceptable success rates (67-86%).

3.3 Displacing waters during docking calculations

Alternatively, it is possible to consider including/removing water molecules during the docking calculation. Huang and Shoichet[18] implemented this approach in DOCK 3.5.54 but did not include a penalty for displacing the water molecule. Verndok *et al*[47] assigned a penalty of ~ 2kJ/mol for the free energy penalty of displacing a water molecule in GOLD. They observe an improvement in docking results by toggling in and out water molecules and using the 2kJ/mol penalty.

AUTODOCK incorporates water molecules and receptor flexibility by combining several structures (and water arrangements) into a single grid[48]. The key feature of this approach is that they use a "clamped" grid where they calculate a weighted average for the potential at a given grid point. The weight for steric clashes (or van der Waals values that are large and positive) are assigned very low weights. This approach drastically lowers the effect of steric clashes due to flexibility or crystallographic waters and permits the docking into a cavity occupied by water in some structures without causing a steric clash. SLIDE[49] also addresses water molecules in docking but addresses it by sing consolv to predict which

waters to keep. Those waters can be displaced during the docking calculation but pay an energy penalty reflected in the ligand score.

FlexX[50] employs a "particle" concept where it a priori determines hydration sites on the receptor surface. "particles" at these sites are then turned on and off during the docking calculation. The docking calculation involves building the ligand by adding small fragments until it is complete. At each building stage, if the ligand clashes with a putative "particle" then the particle it is removed. If not it is kept until the next iteration. Using this scheme, the authors observed a nearly neglible effect on the pose RMS but also report an average rank improvement of 10.2 at a 2.0 A threshold. That is, the enrichment improved which key at the virtual screening stage.

FITTED[51-52] switches on/off water molecules using a scaling factor that goes from 0 to 1 depending on the distances between the ligand and the water molecule in question. Their comparative study with other docking program shows a modest improvement in docked poses versus results obtained after removing all water molecules[53].

Glide incorporates water molecules in their docking to address solvation after ligands are docked. Water probes around the docked poses identify hydration sites. Statistics about howmany hydrogen bonds exist to polar, and non-polar groups help pinpoint which polar groups are essentially disolvated and which waters are in hydrophobic cavities. A penalty is applied if either of these two metrics

Solvation in lead optimization

Water displacement at the lead optimization stage warrants more scrutiny given that compounds at this stage typically require in-house synthesis, may require an entirely new synthetic route and a whole series may be synthesized to explore optimal interactions. Furthermore, there may be differences in water networks between the lead-protein complex the apo structure and it will be those weakly bound waters that will likely release the most free energy when displaced.

Watermap[54-56] is a protocol that uses inhomogeneous solvation theory to calculate the free energy cost of transferring a water molecule from a protein hydration site to bulk solvent. The authors make a compelling case that displacing water molecules makes a significant contribution to ligand binding. Clustering analysis of water molecules around a protein in a molecular dynamics simulation yields the location of hydration sites. An inhomogeneous solvation theory analysis yields the free energy of water molecules at each hydration site. Given a ligand pose, a functional computes how much free energy is gained (or lost) by displacing the waters that occupied the ligand volume. Abel *et al*[55] report a high correlation with relative binding free energies for a series of factor Xa inhibitors (R^2=0.81).

Guimaraes et al[57] went a step further and suggested watermap could help address limitations of MM-GB/SA. For their series of CKD2 and factor Xa inhibitors, they were able to observe a modest improvement in correlation with experiment when the two approaches were combined. R^2 improved from 0.75 to 0.82 when a watermap correction was added to MMGBSA for the factor Xa series. In this pacticular implementation of MMGBSA rescoring the GB protein solvation change upon binding was

removed.   A similar improvement was observed for the CDK2 series; R^2 increased from 0.65 to 0.73. While the effect is modest, the authors point out that the methods should complement each other as they capture complementary aspects of binding.

The most careful way to calculate the free energy cost of water displacement is via free energy calculations[58-60]. Pan et al[61] employed a grand canonical ensemble simulation to guide the design of inhibitors that displace a water molecule and inhibit HIV Tat and human PCAF association. displacement of water by an inhibitor An example is the work by Yu and Rick where they calculated the binding affinity of water molecules in the DNA gyrase binding site via thermodynamic integration[62]. This was an interesting case where a natural mutation introduces a water molecule into the ligand/protein interface that lowers the overall binding affinity by 2 kcal/mol.   More recently Michel et al[42] used a monte-carlo free energy perturbation calculate the net effect of displacing water molecules by optimizing a ligand molecule in the binding sites of p38a MAP kinase, EGFT kinase and scytalone dehydratase. Their calculations highlight some of the sampling issues involved in free energy calculations such as hysteresis and, in a water in scytalone dehydratase, the trapping of water as a ligand modification "grows" into a water cavity. The result was an overestimation of the relative binding free energy of the modified ligand and a realization that careful analysis is necessary for accurate results.

Barilla et al[63] performed monte carlo free energy perturbation calculations coupled with replica exchange[64-65] to calculate the binding energy of 54 water molecules in HIV-1 protease, neuraminidase, trypsin and FxA ligand/protein complexes. Their data showed a statistically significant lower binding free energy for waters known to be displaced by ligands versus those not displaced (-6.2 kcal/mol vs -3.7 kcal/mol).

Other approaches for free energy calculations using water are to combine explicit and implicit solvent models.

In some cases free energy calculations using implicit solvent (MMGBSA) can yield comparable results to explicit solvent calculations[66].

EXPERT OPINION

Alternatively, it is possible to imagine cases where it may be advantageous to introduce a water molecule to mediate a ligand/water contact.

The effect of removing waters on enrichment will depend on the databased used. (the decoys used) and how they fit in the "dry" binding site versus the fully solvated site. In the end, however, this will matter!

Acknowledgements

References

(1)   Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. **1997**, *72*, 1047-1069.
(2)   Lu, Y.; Wang, R.; Yang, C.-Y.; Wang, S. *Journal of Chemical Information and Modeling* **2007**, *47*, 668-675.
(3)   Bashford, D.; Case, D. A. *Annual Review of Physical Chemistry* **2000**, *51*, 129-152.
(4)   Ghosh, A.; Rapp, C. S.; Friesner, R. A. *The Journal of Physical Chemistry B* **1998**, *102*, 10983-10990.
(5)   Tsui, V.; Case, D. A. *Biopolymers* **2000**, *56*, 275-291.
(6)   Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A. *The Journal of Physical Chemistry B* **1997**, *101*, 1190-1197.
(7)   Mark, P.; Nilsson, L. *The Journal of Physical Chemistry A* **2001**, *105*, 9954-9960.
(8)   Mahoney, M. W.; Jorgensen, W. L. *The Journal of Chemical Physics* **2000**, *112*, 8910-8922.
(9)   Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *The Journal of Chemical Physics* **2004**, *120*, 9665-9678.
(10)   Amira, S.; Spångberg, D.; Hermansson, K. *Chemical Physics* **2004**, *303*, 327-334.
(11)   Paesani, F.; Voth, G. A. *The Journal of Physical Chemistry B* **2009**, *113*, 5702-5719.
(12)   Guillot, B. *Journal of Molecular Liquids* **2002**, *101*, 219-260.
(13)   Dunitz, J. D. *Science* **1994**, *264*, 670-.
(14)   Mikol, V.; Papageorgiou, C.; Borer, X. *Journal of Medicinal Chemistry* **1995**, *38*, 3361-3367.
(15)   Griffiths, D. J. *Introduction to Electrodynamics*; Prentice Hall, Inc: Upper Saddle River, NJ, 1999.
(16)   Dill, K. A.; Bromberg, S. *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology*; Garland Science: New York, New York, 2003.
(17)   Born, M. *Zeitschrift für Physik A Hadrons and Nuclei* **1920**, *1*, 45-48.
(18)   Huang, N.; Shoichet, B. K. *Journal of Medicinal Chemistry* **2009**, *51*, 4862-4865.
(19)   Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. *Proteins: Structure, Function, and Bioinformatics* **2002**, *49*, 457-471.
(20)   García-Sosa, A.; Mancera, R. *Journal of Molecular Modeling* **2006**, *12*, 422-431.
(21)   Roberts, B. C.; Mancera, R. L. *Journal of Chemical Information and Modeling* **2008**, *48*, 397-408.
(22)   Thilagavathi, R.; Mancera, R. L. *Journal of Chemical Information and Modeling* **2010**, *50*, 415-421.
(23)   de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. E. *Journal of Medicinal Chemistry* **2005**, *48*, 2308-2318.
(24)   Hritz, J.; de Ruiter, A.; Oostenbrink, C. *Journal of Medicinal Chemistry* **2008**, *51*, 7469-7477.

(25)   Santos, R.; Hritz, J.; Oostenbrink, C. *Journal of Chemical Information and Modeling* **2009**, *50*, 146-154.

(26)   Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. *Proteins: Structure, Function, and Bioinformatics* **1999**, *34*, 4-16.

(27)   Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. *Journal of Molecular Biology* **2008**, *377*, 914-934.

(28)   Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. *Journal of Computational Chemistry* **2003**, *24*, 1637-1656.

(29)   Lee, M. R.; Sun, Y. *Journal of Chemical Theory and Computation* **2007**, *3*, 1106-1119.

(30)   Lyne, P. D.; Lamb, M. L.; Saeh, J. C. *Journal of Medicinal Chemistry* **2006**, *49*, 4805-4808.

(31)   Guimarães, C. R. W.; Cardozo, M. *Journal of Chemical Information and Modeling* **2008**, *48*, 958-970.

(32)   Bottoms, C. A.; White, T. A.; Tanner, J. J. *Proteins: Structure, Function and Bioinformatics* **2006**, *64*, 404-421.

(33)   Zhang, X. J.; Matthews, B. W. *Protein Science* **1994**, *3*, 1031-1039.

(34)   Prasad, B. V. L. S.; Suguna, K. *Acta Crystallographica Section D* **2002**, *58*, 250-259.

(35)   Bottoms, C. A.; Smith, P. E.; Tanner, J. J. *Protein Science* **2002**, *11*, 2125-2137.

(36)   Sreenivasan, U.; Axelsen, P. H. *Biochemistry* **1992**, *31*, 12785-12791.

(37)   Loris, R.; Stas, P. P.; Wyns, L. *Journal of Biological Chemistry* **1994**, *269*, 26722-26733.

(38)   Ehrlich, L.; Reczko, M.; Bohr, H.; Wade, R. *Protein Eng.* **1998**, *11*, 11-19.

(39)   Poornima, C. S.; Dean, P. M. *Journal of Computer-Aided Molecular Design* **1995**, *9*, 513-520.

(40)   Goodford, P. J. *Journal of Medicinal Chemistry* **1985**, *28*, 849-857.

(41)   Cozzini, P.; Fornabaio, M.; Mozzarelli, A.; Spyrakis, F.; Kellogg, G. E.; Abraham, D. J. *International Journal of Quantum Chemistry* **2006**, *106*, 647-651.

(42)   Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *The Journal of Physical Chemistry B* **2009**, *113*, 13337-13346.

(43)   Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. *Journal of Molecular Biology* **1997**, *265*, 445-464.

(44)   Amadasi, A.; Surface, J. A.; Spyrakis, F.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. *Journal of Medicinal Chemistry* **2008**, *51*, 1063-1067.

(45)   Amadasi, A.; Spyrakis, F.; Cozzini, P.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. *Journal of Molecular Biology* **2006**, *358*, 289-309.

(46)   García-Sosa, A. T.; Mancera, R. L.; Dean, P. M. *Journal of Molecular Modeling* **2003**, *9*, 172-182.

(47)   Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. *Journal of Medicinal Chemistry* **2005**, *48*, 6504-6515.

(48)   Österberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. *Proteins: Structure, Function, and Bioinformatics* **2002**, *46*, 34-40.

(49)   Schnecke, V.; Kuhn, L. *Perspectives in Drug Discovery and Design* **2000**, *20*, 171-190.

(50)   Rarey, M.; Kramer, B.; Lengauer, T. *Proteins: Structure, Function, and Bioinformatics* **1999**, *34*, 17-28.

(51)   Corbeil, C. R.; Englebienne, P.; Yannopoulos, C. G.; Chan, L.; Das, S. K.; Bilimoria, D.; L'Heureux, L.; Moitessier, N. *Journal of Chemical Information and Modeling* **2008**, *48*, 902-909.

(52)   Corbeil, C. R.; Englebienne, P.; Moitessier, N. *Journal of Chemical Information and Modeling* **2007**, *47*, 435-449.

(53)   Corbeil, C. R.; Moitessier, N. *Journal of Chemical Information and Modeling* **2009**, *49*, 997-1009.

(54)   Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. *Proceedings of the National Academy of Sciences* **2007**, *104*, 808-813.

(55)   Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *Journal of the American Chemical Society* **2008**, *130*, 2817-2831.

(56)   Beuming, T.; Farid, R.; Sherman, W. *Protein Science* **2009**, *18*, 1609-1619.

(57)   Guimarães, C. R. W.; Mathiowetz, A. M. *Journal of Chemical Information and Modeling* **2010**, *50*, 547-559.

(58)   Chipot, C.; Pearlman, D. A. *Molecular Simulation* **2002**, *28*, 1-12.

(59)   Simonson, T.; Archontis, G.; Karplus, M. *Accounts of Chemical Research* **2002**, *35*, 430-437.

(60)   Kollman, P. *Chemical Reviews* **1993**, *93*, 2395-2417.

(61)   Pan, C.; Mezei, M.; Mujtaba, S.; Muller, M.; Zeng, L.; Li, J.; Wang, Z.; Zhou, M.-M. *Journal of Medicinal Chemistry* **2007**, *50*, 2285-2288.

(62)   Yu, H.; Rick, S. W. *Journal of the American Chemical Society* **2009**, *131*, 6608-6613.

(63)   Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. *Journal of the American Chemical Society* **2007**, *129*, 2577-2587.

(64)   Nagasima, T.; Sugita, Y.; Mitsutake, A.; Okamoto, Y. *Computer Physics Communications* **2002**, *146*, 69-76.

(65)   Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **2000**, *329*, 261-270.

(66)   Michel, J.; Verdonk, M. L.; Essex, J. W. *Journal of Medicinal Chemistry* **2006**, *49*, 7427-7439.